

PCTORGANISATION MONDIALE DE LA PROPRIÉTÉ INTELLECTUELLE
Bureau international

DEMANDE INTERNATIONALE PUBLIÉE EN VERTU DU TRAITE DE COOPERATION EN MATIÈRE DE BREVETS (PCT)

| | | |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| (51) Classification internationale des brevets ⁶ : C12Q 1/68 | A2 | (11) Numéro de publication internationale: WO 97/30178 (43) Date de publication internationale: 21 août 1997 (21.08.97) |
| (21) Numéro de la demande internationale: PCT/FR97/00297 (22) Date de dépôt international: 17 février 1997 (17.02.97) (30) Données relatives à la priorité: 96/01864 15 février 1996 (15.02.96) FR (71) Déposant (pour tous les Etats désignés sauf US): FONDATION JEAN DAUSSET-CEPH [FR/FR]; 27, rue Juliette-Dodu, F-75010 Paris (FR). (72) Inventeurs; et (75) Inventeurs/Déposants (US seulement): NERI, Christian [FR/FR]; 5, rue des Reculettes, F-75013 Paris (FR); CANN, Howard, M. [FR/FR]; 57, rue de Lille, F-75007 Paris (FR); COHEN, Daniel [FR/FR]; 5, avenue Odette, F-94120 Fontenay-sous-Bois (FR). (74) Mandataire: MARTIN, Jean-Jacques; Cabinet Régimbeau, 26, avenue Kléber, F-75116 Paris (FR). | | (81) Etats désignés: CA, JP, US, brevet européen (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE). Publiée <i>Sans rapport de recherche internationale, sera republiée dès réception de ce rapport.</i> |
| (54) Title: DIAGNOSING TRINUCLEOTIDE REPEAT DISEASES AND GENES INVOLVED THEREIN (54) Titre: DIAGNOSTIC DES MALADIES A REPETITION TRINUCLEOTIDIQUE ET GENES IMPLIQUES DANS CES MALADIES (57) Abstract <p>A transcribed DNA sequence with a high level of CAG or CTG repeat codons corresponding to sequences A to I on the table, and alleles and complementary sequences thereof, are disclosed. The sequences are particularly useful for diagnosing trinucleotide repeat diseases.</p> (57) Abrégé <p>La présente invention concerne une séquence d'ADN transcrit et riche en triplet répété CAG ou CTG correspondant aux séquences A à I selon le tableau ainsi que leurs allèles et les séquences complémentaires. Ces séquences sont utiles notamment dans le diagnostic des maladies à répétition trinucéotidique.</p> | | |

UNIQUEMENT A TITRE D'INFORMATION

Codes utilisés pour identifier les Etats parties au PCT, sur les pages de couverture des brochures publiant des demandes internationales en vertu du PCT.

| | | | | | |
|----|---------------------------|----|-----------------------------------------------|----|-----------------------|
| AT | Autriche | GB | Royaume-Uni | MW | Malawi |
| AT | Autriche | GE | Géorgie | MX | Mexique |
| AU | Australie | GN | Guinée | NE | Niger |
| BB | Barbade | GR | Grèce | NL | Pays-Bas |
| BE | Belgique | HU | Hongrie | NO | Norvège |
| BF | Burkina Faso | IE | Irlande | NZ | Nouvelle-Zélande |
| BG | Bulgarie | IT | Italie | PL | Pologne |
| BJ | Bénin | JP | Japon | PT | Portugal |
| BR | Bresil | KE | Kenya | RO | Roumanie |
| BY | Belarus | KG | Kirghizistan | RU | Fédération de Russie |
| CA | Canada | KF | République populaire démocratique de Corée | SD | Soudan |
| CF | République centrafricaine | KR | République de Corée | SE | Suède |
| CG | Congo | KZ | Kazakhstan | SG | Singapour |
| CH | Suisse | LI | Liechtenstein | SI | Slovenie |
| CI | Côte d'Ivoire | LK | Sri Lanka | SK | Slovaquie |
| CM | Cameroon | LR | Libéria | SN | Sénégal |
| CN | Chine | LT | Lituanie | SZ | Swaziland |
| CS | Tchécoslovaquie | LU | Luxembourg | TD | Tchad |
| CZ | République tchèque | LV | Lettonie | TG | Togo |
| DE | Allemagne | MC | Monaco | TJ | Tadjikistan |
| DK | Danemark | MD | République de Moldova | TT | Trinité-et-Tobago |
| EE | Estonie | MG | Madagascar | UA | Ukraine |
| ES | Espagne | ML | Mali | UG | Ouganda |
| FI | Finlande | MN | Mongolie | US | Etats-Unis d'Amérique |
| FR | France | MR | Mauritanie | UZ | Ouzbékistan |
| GA | Gabon | | | VN | Viet Nam |

DIAGNOSTIC DES MALADIES A REPETITION TRINUCLEOTIDIQUE
ET GENES IMPLIQUES DANS CES MALADIES

La présente invention concerne la mise en évidence de gènes humains comportant des séquences à triplet répété CAG ou CTG, ainsi que
5 l'utilisation de ces gènes dans le diagnostic et l'éventuel traitement de certaines maladies neurologiques héréditaires.

L'expansion de séquences à triplet répété CAG ou CTG hautement polymorphiques et instables constitue une mutation impliquée dans au moins 6 maladies neurologiques héréditaires humaines (appelées
10 "maladies à triplet répété"), notamment l'atrophie musculaire spinobulbaire (SBMA), la dystrophie myotonique (MD), l'ataxie cérébro-spinale (SCAS) 1 et 3, l'atrophie dentato-rubro-pallidoluisienne (DRPLA) et la maladie d'Huntington (HD) (1).

L'expansion des CAG/CTG (appelée également mutation
15 dynamique) dans chacune de ces maladies est associée à un phénomène d'anticipation, la taille des répétitions étant souvent corrélée de façon inverse avec l'âge de la survenue et/ou de la sévérité des symptômes.

L'expansion des répétitions CAG/CTG peut apparaître dans les régions non codantes (MD) ou codantes (par exemple SBMA et HD) des
20 transcrits. Ces transcrits sont parfois exprimés dans les tissus autres que le cerveau et lorsqu'ils sont traduits conduisent à des produits de gènes plus grands portant un domaine polyglutamine anormalement allongé (2 à 4).

La mise en évidence d'expansion CAG/CTG dans l'ADN génomique de patients et/ou l'anticipation dans les familles à risque a
25 suggéré que les mutations dynamiques sont impliquées dans SCA 2, SCA 4 et SCA 5, l'ataxie cérébrale dominante autosomale (ADCA de type 2) et la forme familiale du désordre affectif bipolaire (BPAD) ou la schizophrénie. L'autisme et la démence familiale qui montrent des variabilités dans l'âge de la survenue ou dans la sévérité des symptômes pourraient également être
30 causés par des mutations dynamiques.

Un assez grand nombre de maladies pourraient également avoir pour origine ou impliquer des mutations dynamiques :

- la maladie de Parkinson,
- les paraplégies spastiques,
- 35 - les ataxies cérébrospinale non répertoriées précédemment,
- les cataractes zonulaires,

- les tremblements incontrôlables,
- les neuropathies amyloïdes familiales,
- les arthrites granulomateuses familiales (maladie de Blau),
- les microsomies hémifaciales,
- 5 - certaines anémies et glaucomes,
- les désordres obsessionnels.

Ce qui précède démontre l'importance considérable des maladies à triplet répété et l'importance de leur diagnostic et de leur traitement.

10 La présente invention repose sur une étude systématique des répétitions CAG/CTG (qui seront ci-après désignées par "[CAG]*n*", *n* étant le nombre de répétitions du triplet), cette étude ayant été réalisée avec des banques d'ADNc humains, celles-ci ayant subi un premier criblage général, puis les séquences retenues ayant été sélectionnées sur la base d'un certain
15 nombre de critères permettant de s'assurer que les séquences en cause appartenaient à de nouveaux gènes humains et pouvaient, avec une grande probabilité, être impliquées dans des maladies à triplet répété. Enfin, les séquences sélectionnées ont fait l'objet d'une étude plus complète destinée à permettre leur localisation.

20 Dans le cadre de la présente invention, deux ensembles d'ADNc provenant de cerveaux humains ont été étudiés pour analyser la présence de répétitions CAG en utilisant l'hybridation d'oligonucléotides sur des membranes à haute densité. Les deux librairies ont été obtenues à partir de ARNm à l'aide d'amorce oligo-dT, l'une des librairies étant constituée de
25 clones d'ADNc de cerveau foetal (FB) et l'autre librairie étant constituée de clones d'ADNc normalisés de cerveau de nouveau-né (NIB).

De plus, les ESTs humains dans les banques de données ont été analysés pour détecter la présence de triplets répétés dans les ADNc de tissus humains autres que le cerveau.

30 De façon générale, la présente invention repose sur la mise en évidence de certaines séquences susceptibles d'être impliquées dans les maladies à triplet répété obtenues dans des conditions d'hybridation permettant de sélectionner des ADNc qui contiennent un nombre de séquences [CAG] supérieur à 9, lesquelles sont plus susceptibles d'être
35 polymorphes dans une population normale.

Les conditions complètes d'analyse et de sélection de ces différentes séquences ne seront pas détaillées ci-après, seuls seront fournis les éléments permettant de les localiser et de les résoudre grâce, notamment, aux amorces PCR qui seront décrites ci-après.

5 Plus particulièrement, la présente invention concerne une séquence d'ADN transcrit riche en triplet répété CAG ou CTG correspondant aux séquences A à I du tableau, ainsi que leurs allèles normaux et mutés et les séquences complémentaires.

10 Par "allèles normaux" on entend désigner les allèles tels qu'ils ont été isolés ou tels qu'il peuvent être isolés de prélèvements d'individus normaux, les "allèles mutés" étant les allèles des gènes portant les séquences [CAG] $_n$ anormalement répétées.

15 Il est important de noter que, bien que certaines de ces séquences soient, totalement ou en partie, publiques, les séquences en cause sont pour la première fois identifiées comme faisant partie d'un gène pouvant présenter une mutation dynamique, lesdits gènes n'ayant, par ailleurs, en eux-mêmes jamais été décrits.

20 Les séquences ainsi mises en évidence présentent, notamment pour sept d'entre elles, un pourcentage d'hétérozygotie (pourcentage HTZ) suffisamment important pour être impliquées très directement dans des maladies à triplet répété. Les séquences E et I qui présentent un très faible taux d'hétérozygotie (HTZ : 0,05) sont moins susceptibles d'être impliquées dans de telles maladies.

25 Bien entendu, comme cela est mentionné précédemment, ces séquences sont identifiées dans le tableau et pourront, si besoin est, être résolues en utilisant les amorces suivantes : SEQ ID 1 à 18, les amorces 1 à 18 correspondant, par paire, aux séquences A à I mentionnées précédemment.

30 Les séquences ainsi mises en évidence peuvent, tout d'abord, être utilisées dans le cadre du diagnostic, et plus exactement d'un pronostic. En effet, comme un grand nombre de maladies ayant un support génétique, la mise en évidence de la présence d'une séquence présentant un nombre de répétitions anormal de triplet [CAG] $_n$ ne peut en elle-même assurer de la survenue de la maladie, mais doit être interprétée en fonction d'un ensemble d'autres informations pour permettre, soit un diagnostic très
35 précoce, soit éventuellement une surveillance spécifique, surtout dans les familles à risque.

Ce diagnostic peut être effectué en comparant la séquence d'ADN selon l'invention du patient avec une séquence normale pour détecter la présence de répétitions trinuécléotidiques supplémentaires.

5 Plus particulièrement, la présente invention concerne un procédé de mise en évidence des risques d'apparition d'une maladie à répétition trinuécléotidique chez un patient, caractérisé en ce qu'on compare la séquence d'ADN selon la revendication 1 dudit patient à une séquence normale pour détecter la présence de répétitions trinuécléotidiques supplémentaires.

10 Bien entendu, il existe un grand nombre de méthodes qui permettent la mise en évidence de triplets surnuméraires par rapport à des séquences normales, par exemple il est possible de mettre en évidence des différences de poids moléculaires en utilisant des gels et une méthode de type RFLP. Mais, les méthodes les plus efficaces consistent à pratiquer
15 préalablement à la mise en évidence des différences une opération d'amplification, par toute méthode appropriée, notamment la méthode dite "PCR", même si d'autres méthodes sont utilisables.

Il n'est pas nécessaire ici de décrire en détail la méthode PCR, celle-ci permet d'amplifier de façon considérable une séquence spécifique
20 comprise entre deux séquences appelées "amorces".

Parmi les amorces utilisables dans le cadre des procédés selon l'invention, il faut citer les amorces des SEQ ID 1 à 18 qui constituent deux par deux des paires d'amorces pour chacune des séquences objet de l'invention.

25 En utilisant les amorces décrites précédemment, on amplifie les séquences selon la présente invention et il est alors possible, par comparaison avec un échantillon normal et/ou étalon, de mettre en évidence la présence des triplets surnuméraires et donc la possibilité de survenue d'une maladie liée à ce type de mutation dynamique.

30 Il est également intéressant de noter que ces maladies à mutation dynamique sont connues pour être d'autant plus sévères et survenir d'autant plus tôt qu'est important le nombre de répétitions. Dans ces conditions, la méthode diagnostic peut permettre, non seulement un bon pronostic de la survenue de la maladie, mais également d'évaluer le moment
35 où cette maladie surviendra et/ou son éventuelle sévérité.

La présente invention concerne également les gènes et leurs allèles qui portent, au moins en parties, ces séquences, lesdits gènes étant, bien entendu, impliqués directement dans la survenue de la maladie.

Les gènes correspondant pourront être exprimés dans des
5 cellules par des moyens connus afin de produire les protéines correspondantes.

Il est possible d'envisager l'utilisation desdites protéines dans certains kits de diagnostic par exemple.

De façon générale, ces protéines étant impliquées dans la
10 maladie, on souhaitera en diminuer la quantité, soit en bloquant leur expression par des méthodes appropriées au niveau des gènes ou des éléments de régulation, soit en les fixant ou en les inactivant, par exemple en utilisant des protéines réceptrices jouant le rôle de leurres. Là encore, ces protéines réceptrices ou inactivées pourront être générées in situ par
15 expression des séquences d'ADN correspondantes.

Il est également possible de prévoir la réalisation d'anticorps monoclonaux correspondant à ces protéines afin d'envisager le blocage
desdites protéines lorsque cela est souhaité, l'ensemble de ces opérations pouvant être réalisé directement in vivo par exemple, en utilisant des
20 techniques de thérapie génique, en particulier en utilisant des vecteurs qui porteront les séquences d'expression des gènes.

On a mis en évidence le fait que les protéines présentant des domaines polyglutamines anormalement étendus avaient des propriétés d'aggrégation anormales, tant entre elles, qu'avec d'autres protéines (7, 8).
25 Les agrégats ainsi créés sont probablement impliqués dans la genèse des maladies à triplet répété, c'est pourquoi il est également possible de prévoir une thérapie dans laquelle les agents thérapeutiques viendraient empêcher l'aggrégation, soit en bloquant la molécule comme décrit précédemment, soit en se fixant à la molécule pour empêcher le rapprochement avec
30 d'autres protéines.

On peut, dans le cadre de la thérapie, prévoir d'utiliser des variants complets ou délétés de ces protéines, normales ou non (quant au domaine [CAG]_n).

Les séquences selon la présente invention peuvent être
35 obtenues grâce aux informations figurant au tableau et aux références qui y sont données et en utilisant les séquences d'amorce qui sont décrites dans les identificateurs de séquence ci-joints.

Les deux bibliothèques utilisées sont :

- une première bibliothèque (5) contenant 60 000 ADNc non normalisés de cerveau fœtal humain (clones FB) (Laboratoire du Dr. Hans Lehrach, Max Plank Institute for Molecular Genetics, Berlin, Allemagne) sous-clonés dans le vecteur p-SPORT-1 (Life Technologies, Inc., Gaithersburg, MA, USA), et
- la seconde bibliothèque contient 40 128 ADNc normalisés de cerveau de nouveau-né (clones NIB) sous-clonés dans un vecteur λ mid (6) et qui est une partie des ressources du consortium IMAGE (Lawrence Livermore Lab., Livermore, CA,) et du programme EST Merck WU (Washington University, St-Louis, LO, USA).

D'autre part, un certain nombre d'autres informations ont été recueillies en utilisant Genbank Database (NCBI Bethesda, MA, USA).

- Les méthodes de sélection et d'analyse qui ont été mises en oeuvre ne font pas en elles-mêmes partie de la présente invention puisque l'invention a essentiellement pour objet les séquences ainsi obtenues et leur utilisation, notamment, dans des méthodes de diagnostic ou des méthodes de traitement thérapeutique.

- Dans la présente analyse, on a retenu, non seulement les séquences polymorphiques [CAG] $_n$ parfaites, mais également les séquences [CAG] $_n$ complexes, c'est-à-dire celles qui sont ponctuées par des insertions de triplets ; en effet cette présence de triplets insérés s'observe notamment dans le cas de SCA1, alors que tel n'est pas le cas pour SBMA, MD et HD.

- L'étude présente a montré que le polymorphisme [CAG] $_n$ apparaît lorsque la séquence [CAG] $_n$ contient plus de 9 copies CAG ($n > 9$) ; c'est pourquoi, dans les sélections effectuées, les conditions de stringence sont assez élevées, ce qui a permis de sélectionner rapidement un grand nombre de séquences [CAG] $_n$ dans lesquelles n est supérieur à 9. D'ailleurs, les séquences [CAG] $_n$ dans lesquelles n est inférieur à 10 et qui ont pu être testées dans le cadre de la présente invention ont montré un polymorphisme nul.

Par contre, il n'est pas possible de relever une corrélation directe entre la longueur des séquences [CAG] $_n$ et le degré de polymorphisme.

D'autre part, les nouveaux clones sélectionnés à partir de bibliothèques NIB sont distincts de ceux sélectionnés dans le groupe FB. Ceci est en accord avec le fait que le cerveau humain exprime un grand nombre de gènes distincts à des stades de développement différents.

- 5 Enfin, des présentes observations il ressort que des séquences [CAG]_n supérieures à 9 sont rares (1 pour 2 200) 1 pour 3 000) dans les ADNc humains représentatifs des régions 3' des ARNm.

- Bien que dans l'état actuel de la présente invention, les séquences qui constituent l'objet de l'invention n'aient pas été reliées
10 directement à des affections neurologiques héréditaires, la présence d'extensions anormales CAG peut être reliée avec le risque de survenue d'une maladie neurologique, en particulier lorsqu'il y a une composante familiale.

- Sauf pour le clone 2.116 qui est localisé sur le chromosome
15 3p14 et peut-être impliqué dans ADCA II, les autres séquences hautement polymorphiques [CAG]_n décrites précédemment ne correspondent à aucun locus connu pour des maladies neurologiques héréditaires. Il a été trouvé que ce clone 2.116 correspondait à un transcrit de 2 227 paires de base hautement exprimé dans les muscles squelettiques et faiblement exprimé
20 dans le cerveau d'après l'analyse Northern Blot. La répétition des séquences CAG polymorphiques est localisée au niveau de la région 3' non traduite de l'ARNm. L'ARNm correspondant au clone 2.116 code pour une protéine putative de 137 acides aminés (nt 127 à 538) ne présentant aucune homologie avec des protéines connues dans Genbank.

- 25 La séquence complète du transcrit du clone 2.116 est représentée par SEQ ID 19.

Néanmoins, il existe un assez grand nombre de maladies neurologiques qui actuellement n'ont pas été localisées et trouveront probablement à être reliées avec les séquences selon la présente invention.

- 30 Parmi les éléments additionnels qui ont été pris en compte pour étudier la pertinence des séquences retenues figure l'existence éventuelle d'un cadre de lecture ouvert (ORF) dans lequel les séquences [CAG]_n codent pour une extension polyglutamine.

- Enfin, la présente étude suggère que la fréquence du
35 polymorphisme [CAG]_n est un événement rare dans les ADNc humains représentatifs des régions 3' des ARNm.

TABLEAU

[CAGin polymorphiques : caractéristiques]

| | Clone | Banque d'origine Nom d'origine du clone d'ADNc | EST | Fréquence d'hétéro- zygotie | Nbre d'allèles (Nbre de répétitions) | YAC(s) positifs | Locali- sation |
|---|---------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------|----------------------|-----------------------------------|--------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------|-----------------------------|
| A | 2.116 hybridation (CAG)20 PARFAIT | "Fetal brain" (Lab. Dr. H. Lehrach, MPI/MolGen) ICRFp507E07266 (Nom RLDB) | aucune | 0,875 | 9 (14-24) | 872h9 STS+ Alu probe >675f12 STS+ Alu target 890d7 STS+ Alu probe 872h9 STS+ Alu probe | 3p14 |
| B | 2.81 hybridation (CTG)19 COMPLEXE | "Fetal brain" (Lab. Dr. H. Lehrach, MPI/MolGen) ICRFp507I05222 (Nom RLDB) | aucune | 0,578 | 3 (13-19) | En cours de test | |
| C | 2.119 hybridation (CAG)28 PARFAIT | "Fetal brain" (Lab. Dr. H. Lehrach, MPI/MolGen) ICRFp507J10268 (Nom RLDB) | aucune | 0,225 | 5 (23-28) | 767b11 STS+ Alu target 988d9 STS+ Alu probe | 4p15 4q28.3 |
| D | 2.46 hybrid. & banques de données (CAG)10 PARFAIT | "Fetal brain" (Lab. Dr. H. Lehrach, MPI/MolGen) ICRFp507E06174 (Nom RLDB) #936206 HFBC27 (Nom Genbank) | L10376 T07007 | 0,675 | 9 (8-18) | 748e10 STS+ Alu probe (chr12) >970d8 STS+ Alu target >970d10 STS+ Alu target 756g5 STS+ Alu target (chr3) | 12q13.3 9p21 |
| E | 2.70 hybridation (CAG)15 COMPLEXE | "Fetal brain" (Lab. Dr. H. Lehrach, MPI/MolGen) ICRFp507K24212 (Nom RLDB) | aucune | 0,05 | 2 (8-9) | 926g5 STS+ Alu probe | 19q13.43 |
| F | 1.8 hybrid. & banques de données (CAG)13 PARFAIT | "Normalized infant brain" (réseau IMAGE) 30262 (Nom IMAGE) | R18580 | 0,90 | 15 (8-25) | 903e4 STS+ Alu target voisin de STS+ YACs utilisé comme Alu probes | 13q13.1-2 |

| | | | | | | | |
|---|------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------|----------------------|-------|----------|---------------------------------------------------------------------------|-----------------------------|
| G | 1.180 banques de données (CAG)12 PARFAIT | "Normalized breast (réseau IMAGE) 153781 (Nom IMAGE) "Normalized placenta Nb2Hp" (réseau IMAGE) 162267 (Nom IMAGE) | R48249 H25944 | 0.675 | 4 (6-12) | 852g6.STS+, Alu probe >884f6.STS+, Alu probe >891b4.STS+, Alu probe | 1q32-q41 |
| H | 1.181 banques de données (CAG)10 PARFAIT | "Normalized fetal liver spleen" (réseau IMAGE) 114128 (Nom IMAGE) | T85390 | 0.575 | 2 (7-10) | aucun | chr3.8,11 (aucun YAC) |
| I | 1.182 banques de données (CAG)12 PARFAIT | "Normalized breast 3NbHBst" (réseau IMAGE) 57444 (Nom IMAGE) | T49359 | 0.05 | 3 (9-12) | aucun | chr5 (aucun YAC) |

Clone : Nom du clone d'ADNc (nom CEPH) - Mode de sélection - Répétition 5'-3'

EST : Homologues de séquence avec les EST dans Genbank : numéro d'accès des ESTs dans Genbank

YAC(s) positif(s) : (Noms du CEPH) pour la séquence testée

STS+/STS- : YAC positif/négatif pour d'autres STS(s)

Alu target : YAC non utilisé comme sonde Alu

Alu probe : YAC utilisé comme sonde Alu

Localisation : Localisation de la séquence testée dans les chromosomes humains (déduction d'après les YACs positifs)

LISTE DE SEQUENCES

(1) INFORMATIONS GENERALES

(i) DEPOSANT :

- 5 (A) NOM : FONDATION JEAN DAUSSET - CENTRE D'ETUDE DU
POLYMORPHISME HUMAIN (CEPH).
(B) RUE : 27 Rue Juliette Dodu
(C) VILLE : PARIS
(D) PAYS : FRANCE
10 (E) CODE POSTAL : 75010

(ii) TITRE DE L'INVENTION : DIAGNOSTIC DES MALADIES A REPETITION
TRINUCLEOTIDIQUE ET GENES IMPLIQUES
DANS CES MALADIES

15

(iii) NOMBRE DE SEQUENCES : 18

(iv) FORME DECHIFFRABLE PAR ORDINATEUR :

- 20 (A) TYPE DE SUPPORT : Floppy disk
(B) ORDINATEUR : MACINTOSH APPLE
(C) SYSTEME D'EXPLOITATION : MAC-OS SYSTEME 7
(D) LOGICIEL : WORDPERFECT

(v) DONNEES DE LA DEMANDE ACTUELLE :

25 NUMERO DE LA DEMANDE :

(2) INFORMATIONS POUR LA SEQ.ID NO 1 :

(i) CARACTERISTIQUES DE LA SEQUENCE:

- 30 (A) LONGUEUR : 15 nucléotides
(B) TYPE : nucléotide
(C) NOMBRE DE BRIN : simple
(D) CONFIGURATION : linéaire

35 (ii) TYPE DE MOLECULE : ADNc

(iii) CARACTERISTIQUES DU CLONE CELLULAIRE PORTEUR DE LA MOLECULE :

(A) TYPE DE CELLULE : bactérie

(B) NOM DU CLONE D'ADNc CHEZ LE DEPOSANT : 2.116

(C) ORIGINE : Max Plank Institute for Molecular Genetics
Laboratory of Dr Hans Lehrach, Berlin, Allemagne.

(D) NOM DU CLONE DANS LA BANQUE D'ORIGINE :ICRFp507E07266

(iv) DESCRIPTION DE LA SEQUENCE :

CCAGCCTCAG GTAGC

15

(2) INFORMATIONS POUR LA SEQ.ID NO 2 :

(i) CARACTERISTIQUES DE LA SEQUENCE:

(A) LONGUEUR : 22 nucléotides

(B) TYPE : nucléotide

(C) NOMBRE DE BRIN : simple

(D) CONFIGURATION : linéaire

(ii) TYPE DE MOLECULE : ADNc

(iii) CARACTERISTIQUES DU CLONE CELLULAIRE PORTEUR DE LA MOLECULE :

(A) TYPE DE CELLULE : bactérie

(B) NOM DU CLONE D'ADNc CHEZ LE DEPOSANT : 2.116

(C) ORIGINE : Max Plank Institute for Molecular Genetics,
Laboratory of Dr Hans Lehrach, Berlin, Allemagne

(D) NOM DU CLONE DANS LA BANQUE D'ORIGINE :ICRFp507E07266

(iv) DESCRIPTION DE LA SEQUENCE :

GTACTGAGGG CTTTTAGAT TC

22

(2) INFORMATIONS POUR LA SEQ ID NO 3 :

(i) CARACTERISTIQUES DE LA SEQUENCE:

5

(A) LONGUEUR : 18 nucléotides

(B) TYPE : nucléotide

(C) NOMBRE DE BRIN : simple

(D) CONFIGURATION : linéaire

(ii) TYPE DE MOLECULE : ADNc

10

(iii) CARACTERISTIQUES DU CLONE CELLULAIRE PORTEUR DE LA MOLECULE:

(A) TYPE DE CELLULE : bactérie

(B) NOM DU CLONE D'ADNc CHEZ LE DEPOSANT : 2.119

15

(C) ORIGINE : Max Plank Institute for Molecular Genetics,

Laboratory of Dr Hans Lehrach, Berlin, Allemagne

(D) NOM DU CLONE DANS LA BANQUE D'ORIGINE :ICRFp507J10268

(iv) DESCRIPTION DE LA SEQUENCE:

20

TCATGCAGCA GAAAACAG

18

(2) INFORMATIONS POUR LA SEQ ID NO 4 :

25

(i) CARACTERISTIQUES DE LA SEQUENCE:

(A) LONGUEUR : 18 nucléotides

(B) TYPE : nucléotide

(C) NOMBRE DE BRIN : simple

(D) CONFIGURATION : linéaire

30

(ii) TYPE DE MOLECULE : ADNc

(iii) CARACTERISTIQUES DU CLONE CELLULAIRE PORTEUR DE LA MOLECULE:

35

- (A) TYPE DE CELLULE : bactérie
(B) NOM DU CLONE D'ADNc CHEZ LE DEPOSANT :
(C) ORIGINE : Max Plank Institute for Molecular Genetics,
Laboratory of Dr Hans Lehrach, Berlin, Allemagne
(D) NOM DU CLONE DANS LA BANQUE D'ORIGINE:ICRFp507J10268

(iv) DESCRIPTION DE LA SEQUENCE :

AAAGGGGAGA CCAATTTG 18

(2) INFORMATIONS POUR LA SEQ.ID NO 5 :

(i) CARACTERISTIQUES DE LA SEQUENCE:

- (A) LONGUEUR : 18 nucléotides
(B) TYPE : nucléotide
(C) NOMBRE DE BRIN : simple
(D) CONFIGURATION : linéaire

(ii) TYPE DE MOLECULE : ADNc

(iii) CARACTERISTIQUES DU CLONE CELLULAIRE PORTEUR DE LA MOLECULE :

- (A) TYPE DE CELLULE : bactérie
(B) NOM DU CLONE D'ADNc CHEZ LE DEPOSANT : 2.70
(C) ORIGINE : Max Plank Institute for Molecular Genetics,
Laboratory of Dr Hans Lehrach, Berlin, Allemagne
(D) NOM DU CLONE DANS LA BANQUE D'ORIGINE:ICRFp507K24212

(iv) DESCRIPTION DE LA SEQUENCE :

GCACAGTCT CACTAAAC 18

(2) INFORMATIONS POUR LA SEQ.ID NO 6 :

(i) CARACTERISTIQUES DE LA SEQUENCE:

- (A) LONGUEUR : 19 nucléotides
(B) TYPE : nucléotide
(C) NOMBRE DE BRIN : simple
(D) CONFIGURATION : linéaire

5

(ii) TYPE DE MOLECULE : ADNc

(iii) CARACTERISTIQUES DU CLONE CELLULAIRE PORTEUR DE LA MOLECULE:

10

- (A) TYPE DE CELLULE : bactérie
(B) NOM DU CLONE D'ADNc CHEZ LE DEPOSANT : 2.70
(C) ORIGINE : Max Plank Institute for Molecular Genetics,
Laboratory of Dr Hans Lehrach, Berlin, Allemagne
(D) NOM DU CLONE DANS LA BANQUE D'ORIGINE:ICRFp507K24212

15

(iv) DESCRIPTION DE LA SEQUENCE :

GGACATTGGC TTCAACTTC

19

20

(2) INFORMATIONS POUR LA SEQ.ID NO 7:

(i) CARACTERISTIQUES DE LA SEQUENCE:

- (A) LONGUEUR : 16 nucléotides
(B) TYPE : nucléotide
(C) NOMBRE DE BRIN : simple
(D) CONFIGURATION : linéaire

25

(ii) TYPE DE MOLECULE : ADNc

30

(iii) CARACTERISTIQUES DU CLONE CELLULAIRE PORTEUR DE LA MOLECULE:

- (A) TYPE DE CELLULE : bactérie
(B) NOM DU CLONE D'ADNc CHEZ LE DEPOSANT : 2.81
(C) ORIGINE : Max Plank Institute for Molecular Genetics,
Laboratory of Dr Hans Lehrach, Berlin, Allemagne
(D) NOM DU CLONE DANS LA BANQUE D'ORIGINE : ICRFp507I0522

35

(iv) DESCRIPTION DE LA SEQUENCE :

CAGGTGCAGC GTCAAA

16

5 (2) INFORMATIONS POUR LA SEQ.ID NO 8 :

(i) CARACTERISTIQUES DE LA SEQUENCE:

(A) LONGUEUR : 16 nucléotides

(B) TYPE : nucléotide

10 (C) NOMBRE DE BRIN : simple

(D) CONFIGURATION : linéaire

(ii) TYPE DE MOLECULE : ADNc

15 (iii) CARACTERISTIQUES DU CLONE CELLULAIRE PORTEUR DE LA MOLECULE :

(A) TYPE DE CELLULE : bactérie

(B) NOM DU CLONE D'ADNc CHEZ LE DEPOSANT : 2.81

20 (C) ORIGINE : Max Plank Institute for Molecular Genetics,
Laboratory of Dr Hans Lehrach, Berlin, Allemagne

(D) NOM DU CLONE DANS LA BANQUE D'ORIGINE :ICRfp507105222

(iv) DESCRIPTION DE LA SEQUENCE :

25 GGAGGAGGTG TCACAG

16

(2) INFORMATIONS POUR LA SEQ.ID NO 9 :

(i) CARACTERISTIQUES DE LA SEQUENCE:

30 (A) LONGUEUR : 19 nucléotides

(B) TYPE : nucléotide

(C) NOMBRE DE BRIN : simple

(D) CONFIGURATION : linéaire

35 (ii) TYPE DE MOLECULE : ADNc

(iii) CARACTERISTIQUES DU CLONE CELLULAIRE PORTEUR DE LA MOLECULE:

(A) TYPE DE CELLULE : bactérie

(B) NOM DU CLONE D'ADNc CHEZ LE DEPOSANT : i.8

(C) ORIGINE : Réseau I.M.A.G.E., Lawrence Livermore National Laboratories, Livermore, CA, USA.

(D) NOM DU CLONE DANS LA BANQUE D'ORIGINE : 30262

(iv) DESCRIPTION DE LA SEQUENCE :

GATAAAAGGA AGGGAAAAG

19

(2) INFORMATIONS POUR LA SEQ.ID NO 10:

(i) CARACTERISTIQUES DE LA SEQUENCE:

(A) LONGUEUR : 17 nucléotides

(B) TYPE : nucléotide

(C) NOMBRE DE BRIN : simple

(D) CONFIGURATION : linéaire

(ii) TYPE DE MOLECULE : ADNc

(iii) CARACTERISTIQUES DU CLONE CELLULAIRE PORTEUR DE LA MOLECULE:

(A) TYPE DE CELLULE : bactérie

(B) NOM DU CLONE D'ADNc CHEZ LE DEPOSANT : i.8

(C) ORIGINE : Réseau I.M.A.G.E., Lawrence Livermore National Laboratories, Livermore, CA, USA.

(D) NOM DU CLONE DANS LA BANQUE D'ORIGINE : 30262

(iv) DESCRIPTION DE LA SEQUENCE :

GCAACACTCA GAAATGG

17

(2) INFORMATIONS POUR LA SEQ ID NO 11 :

(i) CARACTERISTIQUES DE LA SEQUENCE:

- 5 (A) LONGUEUR : 19 nucléotides
(B) TYPE : nucléotide
(C) NOMBRE DE BRIN : simple
(D) CONFIGURATION : linéaire

(ii) TYPE DE MOLECULE : ADNc

10

(iii) CARACTERISTIQUES DU CLONE CELLULAIRE PORTEUR DE LA MOLECULE :

- 15 (A) TYPE DE CELLULE : bactérie
(B) NOM DU CLONE D'ADNc CHEZ LE DEPOSANT : i.180
(C) ORIGINE : Réseau I.M.A.G.E., Lawrence Livermore National
Laboratories, Livermore, CA, USA.
(D) NOM DU CLONE DANS LA BANQUE D'ORIGINE:153781 ou
162267

20 (iv) DESCRIPTION DE LA SEQUENCE :

AAGTCAGAGT TACTCTTGC

19

(2) INFORMATIONS POUR LA SEQ ID NO 12 :

25

(i) CARACTERISTIQUES DE LA SEQUENCE:

- 30 (A) LONGUEUR : 17 nucléotides
(B) TYPE : nucléotide
(C) NOMBRE DE BRIN : simple
(D) CONFIGURATION : linéaire

(ii) TYPE DE MOLECULE : ADNc

(iii) CARACTERISTIQUES DU CLONE CELLULAIRE PORTEUR DE LA MOLECULE:

(A) TYPE DE CELLULE : bactérie

(B) NOM DU CLONE D'ADNc CHEZ LE DEPOSANT : i.180

(C) ORIGINE : Réseau I.M.A.G.E., Lawrence Livermore National Laboratories, Livermore, CA, USA.

(D) NOM DU CLONE DANS LA BANQUE D'ORIGINE : 153781 ou 162267

(iv) DESCRIPTION DE LA SEQUENCE :

GAGTGAAGTT CAGGAGG

17

(2) INFORMATIONS POUR LA SEQ ID NO 13 :

(i) CARACTERISTIQUES DE LA SEQUENCE:

(A) LONGUEUR : 19 nucléotides

(B) TYPE : nucléotide

(C) NOMBRE DE BRIN : simple

(D) CONFIGURATION : linéaire

(ii) TYPE DE MOLECULE : ADNc

(iii) CARACTERISTIQUES DU CLONE CELLULAIRE PORTEUR DE LA MOLECULE:

(A) TYPE DE CELLULE : bactérie

(B) NOM DU CLONE D'ADNc CHEZ LE DEPOSANT : i.181

(C) ORIGINE : Réseau I.M.A.G.E., Lawrence Livermore National Laboratories, Livermore, CA, USA.

(D) NOM DU CLONE DANS LA BANQUE D'ORIGINE : 114128

(iv) DESCRIPTION DE LA SEQUENCE :

GGACAAAGCT ACATGTCAG

19

(2) INFORMATIONS POUR LA SEQ.ID NO 14 :

(i) CARACTERISTIQUES DE LA SEQUENCE:

5

(A) LONGUEUR : 16 nucléotides

(B) TYPE : nucléotide

(C) NOMBRE DE BRIN : simple

(D) CONFIGURATION : linéaire

(ii) TYPE DE MOLECULE : ADNc

10

(iii) CARACTERISTIQUES DU CLONE CELLULAIRE PORTEUR DE LA MOLECULE :

(A) TYPE DE CELLULE : bactérie

(B) NOM DU CLONE D'ADNc CHEZ LE DEPOSANT : i.181

15

(C) ORIGINE : Réseau I.M.A.G.E., Lawrence Livermore National Laboratories, Livermore, CA, USA.

(D) NOM DU CLONE DANS LA BANQUE D'ORIGINE : 114128

(iv) DESCRIPTION DE LA SEQUENCE :

20

GGTGAGTGTG CTTCTG

16

(2) INFORMATIONS POUR LA SEQ.ID NO 15 :

25

(i) CARACTERISTIQUES DE LA SEQUENCE:

(A) LONGUEUR : 15 nucléotides

(B) TYPE : nucléotide

(C) NOMBRE DE BRIN : simple

(D) CONFIGURATION : linéaire

30

(ii) TYPE DE MOLECULE : ADNc

(iii) CARACTERISTIQUES DU CLONE CELLULAIRE PORTEUR DE LA MOLECULE :

35

(A) TYPE DE CELLULE : bactérie

(B) NOM DU CLONE D'ADNc CHEZ LE DEPOSANT : i.182

(C) ORIGINE : Réseau I.M.A.G.E., Lawrence Livermore National
Laboratories, Livermore, CA, USA.

(D) NOM DU CLONE DANS LA BANQUE D'ORIGINE : 67444

5 (iv) DESCRIPTION DE LA SEQUENCE :

GGGCTAAGGG GAAAG

15

(2) INFORMATIONS POUR LA SEQ ID NO 16 :

10

(i) CARACTERISTIQUES DE LA SEQUENCE:

(A) LONGUEUR : 15 nucléotides

(B) TYPE : nucléotide

(C) NOMBRE DE BRIN : simple

15

(D) CONFIGURATION : linéaire

(ii) TYPE DE MOLECULE : ADNc

(iii) CARACTERISTIQUES DU CLONE CELLULAIRE PORTEUR DE LA
MOLECULE :

20

(A) TYPE DE CELLULE : bactérie

(B) NOM DU CLONE D'ADNc CHEZ LE DEPOSANT : i.182

(C) ORIGINE : Réseau I.M.A.G.E., Lawrence Livermore National
Laboratories, Livermore, CA, USA.

25

(D) NOM DU CLONE DANS LA BANQUE D'ORIGINE : 67444

(iv) DESCRIPTION DE LA SEQUENCE :

CTTGGTGGGC AAGTG

15

30

(2) INFORMATIONS POUR LA SEQ ID NO 17 :

(i) CARACTERISTIQUES DE LA SEQUENCE:

(A) LONGUEUR : 18 nucléotides

35

(B) TYPE : nucléotide

(C) NOMBRE DE BRIN : simple

(D) CONFIGURATION : linéaire

(ii) TYPE DE MOLECULE : ADNc

(iii) CARACTERISTIQUES DU CLONE CELLULAIRE PORTEUR DE LA MOLECULE :

5

(A) TYPE DE CELLULE : bactérie

(B) NOM DU CLONE D'ADNc CHEZ LE DEPOSANT : 2.46

(C) ORIGINE : Max Plank Institute for molecular Genetics,
Laboratory of Dr Hans Lehrach, Berlin, Allemagne.

(D) NOM DU CLONE DANS LA BANQUE D'ORIGINE :ICRFp507E06174

10

(iv) DESCRIPTION DE LA SEQUENCE :

TTTTTACTCG CGGCGGTG

18

15 (2) INFORMATIONS POUR LA SEQ.ID NO 18 :

(i) CARACTERISTIQUES DE LA SEQUENCE:

(A) LONGUEUR : 25 nucléotides

(B) TYPE : nucléotide

20

(C) NOMBRE DE BRIN : simple

(D) CONFIGURATION : linéaire

(ii) TYPE DE MOLECULE : ADNc

25

(iii) CARACTERISTIQUES DU CLONE CELLULAIRE PORTEUR DE LA MOLECULE :

(A) TYPE DE CELLULE : bactérie

(B) NOM DU CLONE D'ADNc CHEZ LE DEPOSANT : 2.46

(C) ORIGINE : Max Plank Institute for molecular Genetics,
Laboratory of Dr Hans Lehrach, Berlin, Allemagne.

30

(D) NOM DU CLONE DANS LA BANQUE D'ORIGINE :ICRFp507E06174

(iv) DESCRIPTION DE LA SEQUENCE :

35

CAGGATAATC AAGAAATGAAG TTAAG

25

REFERENCES

1. Whilhem, P.J. Dynamic mutations hit double figures. *Nature Genet.* 8, 213-215 (1994).
2. Trottier, Y. et al. Cellular localization of the Huntington's disease protein and discrimination of the normal and mutant form. *Nature Genet.* 10, 104-110 (1995).
3. Servadio, A. et al. Expression analysis of the ataxin-1 protein in tissues from normal and spinocerebellar ataxia type 1 individuals. *Nature Genet.* 10, 94-98 (1995).
4. Yazawa, I. et al. Abnormal gene product identified in hereditary dentato-rubralpallidoluysian atrophy (DRPLA) brain. *Nature Genet.* 10, 99-103 (1995).
5. Meier-Ewert, S., Maier, E., Ahmadi, A., Curtis, J. and Lebrach, H. An Automated approach to generating expressed sequence catalogues. *Nature* 361, 375 (1993).
6. Soares, M.B., Bonaldo, M.F., Jelene, P., Su, L., Lawton, L. & Efstratiadis, A. Construction and characterization of a normalized cDNA library. *Proc. Natl. Acad. Sci. USA.* 91, 9228-9232 (1994).
7. Haad, D., Perry, M.J. and Haynes, L. Cellular transglutaminases in neural development. *Int. J. Devl Neuroscience* 11, 709-720 (1993).
8. Stott, K., Blackburn, J.M., Butler, P.J.G. & Perutz, M. Incorporation of glutamine repeats makes protein oligomerize. Implications for neuro-degenerative diseases. *Proc. Natl. Acad. Sci. USA.* 92, 6509-6513 (1995).

REVENDICATIONS

- 1) Séquence d'ADN transcrit riche en triplet répété CAG ou CTG correspondant aux séquences A à I selon le tableau ainsi que leurs allèles normaux et mutés et les séquences complémentaires.
- 2) Gène humain et ses allèles, caractérisé en ce qu'il porte au moins en partie une séquence selon la revendication 1.
- 3) Procédé de mise en évidence des risques d'apparition d'une maladie à répétition trinuécléotidique chez un patient, caractérisé en ce qu'on compare la séquence d'ADN selon la revendication 1 dudit patient à une séquence normale pour détecter la présence de répétitions trinuécléotidiques supplémentaires.
- 4) Procédé selon la revendication 3, caractérisé en ce qu'on effectue une amplification de tout ou partie des séquences selon la revendication 1 et que l'on met en évidence dans le produit d'amplification la présence de répétitions trinuécléotidiques supplémentaires.
- 5) Procédé selon la revendication 4, caractérisé en ce que l'amplification des séquences est effectuée avec les amorces décrites SEQ ID 1 à 18 correspondant aux séquences A à I respectivement.
- 6) Cellule transformée exprimant tout ou partie d'un gène selon la revendication 2.
- 7) Protéine obtenue à partir d'une cellule transformée selon la revendication 6.
- 8) Protéine interagissant avec une protéine selon la revendication 7.
- 9) Anticorps monoclonal correspondant à une protéine selon l'une des revendications 7 ou 8.
- 10) Vecteur exprimant une protéine selon l'une des revendications 7 ou 8.
- 11) Utilisation thérapeutique d'une protéine selon l'une des revendications 7 ou 8, d'un anticorps selon la revendication 9 ou d'un vecteur selon la revendication 10.